

Conditional independence relations and log-linear models for random permutations

Villő Csiszár *

February 2, 2008

Abstract

We propose a new class of models for random permutations, which we call log-linear models, by the analogy with log-linear models used in the analysis of contingency tables. As a special case, we study the family of all Luce-decomposable distributions, and the family of those random permutations, for which the distribution of both the permutation and its inverse is Luce-decomposable. We show that these latter models can be described by conditional independence relations. We calculate the number of free parameters in these models, and describe an iterative algorithm for maximum likelihood estimation, which enables us to test if a set of data satisfies the conditional independence relations or not.

AMS 2000 Subject classifications: 62H05, 62E10, 60E05, 62-07.

Keywords: ranking models, random permutations, log-linear models, conditional independence, decomposability.

1 Introduction

There are three slightly different situations, in which permutation-valued data may turn up.

(i) The permutation describes a *pairing*, i.e. a one to one correspondance, between two sets of cardinality n , whose elements are labelled with the numbers $1, \dots, n$. An example is a pairing of boys and girls for a dance. If the first set is the boys and the second set is the girls, then a pairing is given by the permutation π , where $\pi(i) = j$ means that boy i dances with girl j .

(ii) The permutation describes a *ranking* of a labelled set of cardinality n , i.e. the ranking from best to worst of labelled alternatives. The ranking is given by π , where $\pi(i) = j$ means that alternative i is ranked j th best. The inverse of the ranking π is the *ordering* π^{-1} , i.e. $\pi^{-1}(i)$ is the label of the alternative ranked i th best.

(iii) The permutation describes a *reordering* of a set of ordered elements. For example, books on a shelf in a library are reordered as readers look into them. Here $\pi(i) = j$ means that the i th item in the original order becomes the j th item after reordering.

*Loránd Eötvös University, Budapest, e-mail: csvillo@gmail.com

Notice that a pairing or a ranking/ordering can be described by a permutation only after a labelling is fixed on the sets.

There is a vast literature of models for random permutations, especially for ranking/ordering data (*ii*). Comprehensive summaries can be found, among others, in Critchlow, Fligner, and Verducci [4] and Marden [17]. The larger classes of models are order statistics models (also called Thurstonian models), distance-based models, paired comparison models, and multistage models.

Our starting point is the concept of Luce-decomposability, also called L -decomposability (we will use the latter name). This property, applied to orderings, was introduced by Critchlow, Fligner and Verducci in [4], motivated by Luce's ranking postulate [15]. This postulate supposes that the ordering of the alternatives is the result of repeated selections of the best alternative from the remaining set of alternatives. That is, for each set C and each alternative $x \in C$, the probability that x is chosen as best from C is given by $p_C(x)$. Given these choice probabilities, the probability of the ordering π is given by

$$p(\pi) = \prod_{k=1}^n p_{C_k}(\pi(k)),$$

where $C_k = \{\pi(k), \dots, \pi(n)\}$ is the set of available alternatives at the k th step. Luce combined this postulate with his choice axiom to develop the Luce model. The choice axiom puts restrictions on the choice probabilities $p_C(x)$. The ranking postulate without the choice axiom produces a multistage model, a general L -decomposable distribution. In other words, a random ordering (or its distribution) is L -decomposable, if its elements are chosen successively, satisfying the following Markov property: in the k th step, the k th element is chosen from the remaining ones, independently of the order of the first $k - 1$ elements.

In this paper, we wish to apply L -decomposability to random pairings, rankings, and reorderings as well. We notice that L -decomposability of a random pairing depends on the labelling of the first set, and L -decomposability of a random ranking depends on the labelling of the alternatives. In fact, the labelling under which L -decomposability is satisfied (if such a labelling exists) can be interpreted as a natural order of the elements or alternatives. If n is relatively small, this natural order may be found by an exhaustive search over all labellings.

In the main part of the paper, we study random permutations Π , for which the distributions of both Π and Π^{-1} are L -decomposable. In this case we say that the distribution of Π (and of Π^{-1}) is bi-decomposable. Bi-decomposability of a random *pairing* implies a natural order on both sets, while bi-decomposability of a random *ranking/ordering* implies a natural order on the set of alternatives. This new concept is perhaps most natural for random pairings, since they possess an obvious symmetry in the two sets whose elements are

paired, that is in Π and Π^{-1} . The model is also attractive in the case of rankings/orderings, since, as we shall show, a general L -decomposable distribution has $2^n(n/2 - 1) + 1$ free parameters, while a general bi-decomposable distribution possesses only $\sum_{k=1}^{n-1} k^2$ parameters. This is still more than the usual number of about n or at most n^2 for most known models, however, it is still very small compared to $n!$.

Another feature of both the L -decomposable and the bi-decomposable families is that they can be characterised by certain conditional independence relations, which we will describe later in detail. Therefore, by fitting these models with the maximum likelihood method, and assessing the goodness of fit with the chi-square test, we can test the hypothesis that the random permutation satisfies certain conditional independence properties.

The paper is organised as follows. Section 2 deals with L -decomposability, and Section 3 contains the main results about bi-decomposable distributions. In Section 3.1, we study general log-linear models for random permutations, which we apply in Section 3.2 to decomposable models, and prove Theorem 1. Section 3.3 treats the problem of maximum likelihood estimation in the models. It is shown that the maximum likelihood estimate is explicit in the L -decomposable model, but in the bi-decomposable model, it can only be obtained by iterative methods. In Section 4, we investigate which models formulated in the literature are L -decomposable or bi-decomposable. In Section 5 we study to what extent the latent order with respect to which the distribution is decomposable can be estimated. Finally, in Section 6, we fit the models to a real dataset, and Section 7 contains the proofs of some lemmas.

2 L -decomposability

For integers $i \leq j$, $\{i : j\}$ denotes the set $\{k : i \leq k \leq j\}$. For any vector $v = (v(1), \dots, v(s))$, we call $v(i)$ the i th element of v . For the set of the i th to j th elements, and for the subvector of the i th to j th elements of v , introduce the notations

$$v\{i : j\} = \{v(i), \dots, v(j)\}, \quad v(i : j) = (v(i), \dots, v(j)), \quad 1 \leq i \leq j \leq s. \quad (1)$$

If $j < i$ then let $v\{i : j\}$ be the empty set. Let S_n stand for the symmetric group of all permutations π of $\{1, \dots, n\}$. We denote a probability distribution on S_n by $p = \{p(\pi) : \pi \in S_n\}$. Denote by $\Pi : \Omega \rightarrow S_n$ a random permutation on a probability space (Ω, \mathcal{A}, P) with distribution p , that is $P(\Pi = \pi) = p(\pi)$. The idea of *L -decomposability* first appears in [4], and was motivated by Luce's ranking postulate [15]. It states that for any k , the value of $\Pi(k + 1)$ depends on $\Pi(1 : k)$ only through $\Pi\{1 : k\}$. Recall that the probability

of a permutation can always be written in the product form

$$P(\Pi = \pi) = \prod_{k=0}^{n-1} P(\Pi(k+1) = \pi(k+1) \mid \Pi(1:k) = \pi(1:k)). \quad (2)$$

L -decomposability means that the conditions $\Pi(1:k) = \pi(1:k)$ can be replaced by the conditions $\Pi\{1:k\} = \pi\{1:k\}$. We formulate this in the following definition, in four different forms. For two permutations, $\pi\sigma$ denotes composition, i.e. $\pi\sigma(i) = \pi(\sigma(i))$.

Definition 1. Let Π be a random permutation with probability distribution p on S_n . Π or p is called L -decomposable, if any of the following are satisfied.

1. For every $2 \leq k \leq n-2$, $\pi \in S_n$ and $\sigma \in S_k$

$$\begin{aligned} P(\Pi(k+1) = \pi(k+1) \mid \Pi(1:k) = \pi(1:k)) &= \\ &= P(\Pi(k+1) = \pi(k+1) \mid \Pi(1:k) = \pi\sigma(1:k)), \end{aligned} \quad (3)$$

if both conditional probabilities are defined.

2. For every $2 \leq k \leq n-2$ and $\pi \in S_n$

$$\begin{aligned} P(\Pi(k+1) = \pi(k+1) \mid \Pi(1:k) = \pi(1:k)) &= \\ &= P(\Pi(k+1) = \pi(k+1) \mid \Pi\{1:k\} = \pi\{1:k\}), \end{aligned} \quad (4)$$

if the lefthandside is defined.

3. The random sets $\Pi\{1:k\}$ form a Markov chain for $k = 1, \dots, n$.
4. There exists a Λ nonnegative function defined on pairs

$$(x, C) : C \subset \{1, \dots, n\}, x \notin C, \quad (5)$$

and c constant, such that for all $\pi \in S_n$

$$p(\pi) = c \prod_{k=0}^{n-1} \Lambda(\pi(k+1), \pi\{1:k\}). \quad (6)$$

Proposition 1. *The four properties in Definition 1 are equivalent.*

This proposition, in slightly different form, can be found in [4], so we omit the straightforward proof. In the first two equivalent forms of the definition, we could formally include $k = 0, 1, n-1$ as well, but equations (3) and (4) are always satisfied for these k -values. It follows that for $n \leq 3$, all distributions are L -decomposable.

The pair (Λ, c) is called an L -decomposition of the distribution p if (6) holds. By (2) and (4), one L -decomposition of the L -decomposable distribution p is given by $c = 1$ and

$$\Lambda(x, C) = P(\Pi(|C| + 1) = x \mid \Pi\{1 : |C|\} = C), \quad (7)$$

if the probability of the condition is positive, otherwise $\Lambda(x, C) = 0$. We call this L -decomposition *canonical*.

The fact that the random sets $\Pi\{1 : k\}$ form a Markov chain is equivalent to the independence of the past and the future, conditional on the present. This means that the first k and last $n - k$ elements of Π are conditionally independent on the condition that the set of the first k elements is given. By the well-known property of Markov chains, this observation generalises to any consecutive partition of the set $\{1, \dots, n\}$.

Any j -tuple $\kappa = (\kappa_1, \dots, \kappa_j)$ with $\kappa_0 = 0 < \kappa_1 < \dots < \kappa_j < n = \kappa_{j+1}$ is a set of *sections*, which define a $\underline{\kappa}$ *consecutive partition* of the set $\{1, \dots, n\}$ into $j + 1$ sets by

$$\underline{\kappa} = (\underline{\kappa}_1, \dots, \underline{\kappa}_{j+1}), \text{ where } \underline{\kappa}_i = \{\kappa_{i-1} + 1 : \kappa_i\}. \quad (8)$$

For the consecutive partition $\underline{\kappa}$ and $\pi \in S_n$, define the vector of unordered marginals

$$\{\pi_{\underline{\kappa}}\} = (\{\pi_{\underline{\kappa}_i}\} : 1 \leq i \leq j + 1), \text{ where } \{\pi_{\underline{\kappa}_i}\} = \pi\{\kappa_{i-1} + 1 : \kappa_i\}. \quad (9)$$

In contrast, $\pi_{\underline{\kappa}_i} = \pi(\kappa_{i-1} + 1 : \kappa_i)$ is an ordered marginal.

L -decomposability means that for any consecutive partition $\underline{\kappa}$ in (8) and any $\pi \in S_n$,

$$P(\Pi = \pi \mid \{\Pi_{\underline{\kappa}}\} = \{\pi_{\underline{\kappa}}\}) = \prod_{i=1}^{j+1} P(\Pi_{\underline{\kappa}_i} = \pi_{\underline{\kappa}_i} \mid \{\Pi_{\underline{\kappa}}\} = \{\pi_{\underline{\kappa}}\}). \quad (10)$$

Thus we have proved the following

Proposition 2. *A random permutation Π is L -decomposable if and only if, for every consecutive partition in (8), the ordered marginals $\Pi_{\underline{\kappa}_i}$, $1 \leq i \leq j + 1$ are conditionally independent, given $\{\Pi_{\underline{\kappa}}\}$, that is, (10) holds.*

In the language of orderings, the unordered marginal $\{\pi_{\underline{\kappa}}\}$ is the partial ordering of shape $\underline{\kappa}$ derived from the ordering π . This gives the set of alternatives receiving the first κ_1 ranks, the set of alternatives receiving the next $\kappa_2 - \kappa_1$ ranks, etc. Thus L -decomposability for orderings means that given the partial ordering of arbitrary shape $\underline{\kappa}$, the orderings within each set of ranks $\underline{\kappa}_i$ are independent.

3 Bi-decomposability

We are interested in random permutations, for which the distribution of both Π and Π^{-1} is L -decomposable. If the distribution of Π is p , then the distribution of Π^{-1} is given by

$p'(\pi) = p(\pi^{-1})$. Thus, Π^{-1} is L -decomposable if and only if

$$p(\pi) = p'(\pi^{-1}) = \prod_{(x,C)} \Lambda'(x, C)^{m_L(\pi^{-1}, (x, C))} = \prod_{(x,C)} \Lambda'(x, C)^{m_{L'}(\pi, (x, C))}, \quad (11)$$

where Λ' is the canonical L -decomposition of p' , and the matrix $M_{L'}$ is derived from the matrix M_L by interchanging each pair of rows corresponding to inverse permutations π and π^{-1} . We call such distributions L' -decomposable, and denote their family by $\mathcal{P}_{L'}$. The family of bi-decomposable distributions will be denoted by $\mathcal{P}_b = \mathcal{P}_L \cap \mathcal{P}_{L'}$.

According to Proposition 2, bi-decomposable random permutations have the property that for every consecutive partition $\underline{\kappa}$ in (8), the ordered marginals $\Pi_{\underline{\kappa}_i}$, $1 \leq i \leq j+1$ are conditionally independent, given $\{\Pi_{\underline{\kappa}}\}$, and the ordered marginals $\Pi_{\underline{\kappa}_i}^{-1}$, $1 \leq i \leq j+1$ are conditionally independent, given $\{\Pi_{\underline{\kappa}}^{-1}\}$. We now show that bi-decomposable random permutations satisfy additional conditional independence statements. Let $\underline{\kappa}$ and $\underline{\lambda}$ be two consecutive partitions. For $\pi \in S_n$, define the ordered marginals

$$\pi_{\underline{\kappa}_i \times \underline{\lambda}_j} = \{(a, \pi(a)) : a \in \underline{\kappa}_i, \pi(a) \in \underline{\lambda}_j\}. \quad (12)$$

We prove the next proposition in Section 7.

Proposition 3. *A random permutation Π is bi-decomposable, if and only if for all pairs of consecutive partitions $\underline{\kappa}$ and $\underline{\lambda}$ of sizes s and t respectively, the ordered marginals $\Pi_{\underline{\kappa}_i \times \underline{\lambda}_j}$ for $1 \leq i \leq s$ and $1 \leq j \leq t$ are conditionally independent, given $\{\Pi_{\underline{\kappa}}\}$ and $\{\Pi_{\underline{\lambda}}^{-1}\}$.*

In the rest of the paper, we focus our attention on strictly positive distributions, i.e. the case when $p(\pi) > 0$ for all $\pi \in S_n$, which can be described as exponential families, or more specifically, as log-linear models. In general, by an exponential family of discrete (strictly positive) distributions $p = (p_1, \dots, p_s)$ we mean the family

$$\left\{ p : \sum_{i=1}^s p_i = 1, \log p \in U \right\}, \quad (13)$$

where U is a t -dimensional linear subspace of \mathbb{R}^s , containing the vector $\mathbf{1} = (1, \dots, 1)^T$. The number of free parameters of the exponential family (13) is $t - 1$. Extending the concept used in the analysis of contingency tables, we may define a log-linear model as an exponential family where the linear subspace U has a generating set consisting of $0 - 1$ vectors. All log-linear models for random permutations appearing in this paper have the additional property that the generating $0 - 1$ vectors of U are indicator vectors of different values of generalised marginals $|\pi_B|$.

Denote by \mathcal{P}_L^+ and $\mathcal{P}_{L'}^+$ the family of strictly positive L -decomposable and L' -decomposable distributions respectively. Then the family of strictly positive bi-decomposable distributions is given by $\mathcal{P}_b^+ = \mathcal{P}_L^+ \cap \mathcal{P}_{L'}^+$. We will show that both \mathcal{P}_L^+ and $\mathcal{P}_{L'}^+$ admit a log-linear

representation with corresponding linear subspaces F and G respectively. It follows that \mathcal{P}_b^+ is also an exponential family with subspace $H = F \cap G$. We will show that \mathcal{P}_b^+ is also a log-linear model, and we will determine the dimension and a basis of H (Theorem 1). This is made possible by the abundance of orthogonality. Two subspaces U and V of a Hilbert space are called orthogonal, if every pair of vectors $u \in U, v \in V$ are orthogonal. The closed subspaces U and V intersect each other orthogonally, if the (orthogonal) projection of U on V equals $U \cap V$, or equivalently, the projection of V on U equals $U \cap V$. Denote the operator of orthogonal projection on U by Pr_U . Another equivalent condition for orthogonal intersection is that the projection operators Pr_U and Pr_V commute. Thus introducing the notation \perp_\cap for orthogonal intersection,

$$U \perp_\cap V \iff Pr_U V = U \cap V \iff Pr_V U = U \cap V \iff Pr_U Pr_V = Pr_V Pr_U. \quad (14)$$

We will show that F and G intersect each other orthogonally, furthermore, we will find an orthogonal decomposition of both F and G into lower dimensional subspaces F_k and G_ℓ , such that each pair of subspaces (F_k, G_ℓ) intersect each other orthogonally. Then it will suffice to determine the dimension and basis of the low dimensional subspaces $F_k \cap G_\ell$. Orthogonal intersection does not appear by coincidence, it is the consequence of conditional independence relations, as we will explain later on. Before carrying out this program in the following subsections, we state the main theorem of this section.

Theorem 1. *The family of positive bi-decomposable distributions is a log-linear model with the number of free parameters equal to*

$$d_n = \sum_{i=1}^{n-1} i^2. \quad (15)$$

The proof of Theorem 1 is given in Section 3.2.

3.1 Partitions of the chessboard

A permutation may be identified with a placement of n rooks on the $n \times n$ chessboard such that they cannot capture each other, i.e. a placement with exactly one rook in each row and in each column. Let us agree that we place the rooks “row-wise”, that is if $\pi(i) = j$, then we place a rook in the j th square of the i th row. π^{-1} can be read “column-wise” from the rook-placement. This identification is helpful in the study of bi-decomposability, because bi-decomposability is a symmetric property in π and π^{-1} , that is in rows and columns of the chessboard.

In this section, we define and study log-linear models for random permutations, whose generators are partitions of the chessboard. More specifically, we require that these partitions be the product of a row-partition and a column-partition. A partition of the set

$\{1 : n\}$ into s disjoint subsets (also called atoms) is given by

$$\mathcal{Z} = (Z_1, \dots, Z_s) : \cup_{i=1}^s Z_i = \{1 : n\}, Z_i \cap Z_j = \emptyset \forall i \neq j. \quad (16)$$

If none of the sets Z_i is empty, we call s the size of the partition. If of two such partitions, one partitions the set of rows, the other the set of columns of the $n \times n$ chessboard, then the result is a product partition of the board.

Definition 2. A partition \mathcal{B} of the $n \times n$ chessboard is a *product partition*, if there exist a partition \mathcal{R} of size r (called row-partition) and a partition \mathcal{C} of size c (called column-partition) of the set $\{1 : n\}$ such that

$$\mathcal{B} = (B_{ij}) : B_{ij} = R_i \times C_j = \{(x, y) : x \in R_i, y \in C_j\}, \quad 1 \leq i \leq r, 1 \leq j \leq c. \quad (17)$$

We denote this by $\mathcal{B} = \mathcal{R} \times \mathcal{C}$.

For any product partition \mathcal{B} of the $n \times n$ chessboard, we define the matrix-valued \mathcal{B} -marginal function $\pi \mapsto |\pi_{\mathcal{B}}|$ on S_n . For a permutation π , this statistic gives the number of rooks falling into each B_{ij} in the rook-placement corresponding to π :

$$|\pi_{\mathcal{B}}| = (t_{ij}), \quad t_{ij} = |\{1 \leq s \leq n : (s, \pi(s)) \in B_{ij}\}|. \quad (18)$$

In other words, if π is a pairing between two labelled sets A and B , then $|\pi_{\mathcal{B}}|$ is the $r \times c$ matrix whose ij th entry is the number of elements of A belonging to R_i , which are paired with an element of B belonging to C_j . The partition \mathcal{B} of the chessboard gives rise to a partition of S_n via the function $\pi \mapsto |\pi_{\mathcal{B}}|$: the permutations π and σ belong to the same atom of this partition if and only if $|\pi_{\mathcal{B}}| = |\sigma_{\mathcal{B}}|$. The subspace of $\mathbb{R}^{n!}$ spanned by the indicator vectors of these atoms will be denoted by $U^{\mathcal{B}}$. Equivalently

$$U^{\mathcal{B}} = \{v \in \mathbb{R}^{n!} : |\pi_{\mathcal{B}}| = |\sigma_{\mathcal{B}}| \Rightarrow v(\pi) = v(\sigma)\}. \quad (19)$$

The vectors $v \in U^{\mathcal{B}}$ are just the functions $\pi \mapsto v(\pi)$ on S_n , which are measurable with respect to the (atomic) σ -algebra with atoms $\{\pi : |\pi_{\mathcal{B}}| = (t_{ij})\}$, where (t_{ij}) takes all possible values. We denote this σ -algebra by $\sigma(\mathcal{B})$.

We will define a log-linear model by a set of product partitions, called the generators of the model. Of course, a similar definition is possible also with generator partitions which are not of product form. For the spanned subspace, we use the notation $\text{Span}(\cdot)$.

Definition 3. Let $\mathcal{B}_1, \dots, \mathcal{B}_s$ be product partitions of the chessboard, and use the simplifying notation $U^{\mathcal{B}_i} = U^i$. We say that p belongs to the *log-linear model generated by these partitions* if

$$\log p(\pi) = \sum_{i=1}^s \theta^i(|\pi_{\mathcal{B}_i}|) \quad \pi \in S_n, \quad (20)$$

where the θ^i functions are arbitrary parameters. Equivalently, we require that

$$\log p \in \text{Span}(U^1, \dots, U^s). \quad (21)$$

We will use the notation $\mathcal{L}(\mathcal{B}_1, \dots, \mathcal{B}_s)$ for this model.

In the rest of this section, we give a sufficient condition, when the intersection of two log-linear models is itself a log-linear model, with directly identifiable generators. The proofs can be found in Section 7. The first lemma describes the relationship between conditional independence and orthogonal intersection.

Lemma 1. *Let (Ω, \mathcal{A}, P) be a probability space, and denote by $L_2(\mathcal{A})$ the Hilbert space of square-integrable random variables on it. For a σ -algebra $\mathcal{D} \subset \mathcal{A}$, denote by $L_2(\mathcal{D})$ the closed linear subspace of $L_2(\mathcal{A})$ consisting of all \mathcal{D} -measurable random variables. Let $\mathcal{D}_1, \mathcal{D}_2 \subset \mathcal{A}$. Then $L_2(\mathcal{D}_1) \perp_{\cap} L_2(\mathcal{D}_2)$ if and only if \mathcal{D}_1 and \mathcal{D}_2 are conditionally independent, given $\mathcal{D}_1 \cap \mathcal{D}_2$.*

There is a partial ordering on the set of partitions. Partition $\mathcal{Z} = (Z_1, \dots, Z_s)$ is finer than $\mathcal{W} = (W_1, \dots, W_t)$ (or \mathcal{W} is coarser than \mathcal{Z}) if for every i there exists a j such that $Z_i \subset W_j$. Denote this by $\mathcal{Z} \succ \mathcal{W}$. Clearly, this implies $U^{\mathcal{Z}} \supset U^{\mathcal{W}}$. By the application of Lemma 1, we get

Lemma 2. *Let $\mathcal{R}' \succ \mathcal{R}$ and $\mathcal{C}' \succ \mathcal{C}$ be partitions of $\{1 : n\}$. Then we have*

$$U^{\mathcal{R} \times \mathcal{C}'} \perp_{\cap} U^{\mathcal{R}' \times \mathcal{C}} \text{ and } U^{\mathcal{R} \times \mathcal{C}'} \cap U^{\mathcal{R}' \times \mathcal{C}} = U^{\mathcal{R} \times \mathcal{C}}. \quad (22)$$

The next two lemmas formulate simple facts from linear algebra, which will be needed in the sequel. We write $U = U_1 \oplus U_2$ for orthogonal decomposition, that is when $U = \text{Span}(U_1, U_2)$ and U_1 and U_2 are orthogonal.

Lemma 3. *Suppose that $U = \text{Span}(U_i : i \in I)$, $V = \text{Span}(V_j : j \in J)$ are two subspaces, and $U_i \perp_{\cap} V_j$ for every pair i, j . Then $U \perp_{\cap} V$, and $U \cap V = \text{Span}(U_i \cap V_j : i \in I, j \in J)$.*

Lemma 4. *Let $U = U_1 \oplus U_2$ and $V = V_1 \oplus V_2$ be two subspaces with orthogonal decompositions. If $U \perp_{\cap} V$, $U_1 \perp_{\cap} V_1$, $U \perp_{\cap} V_1$, and $U_1 \perp_{\cap} V$ hold, then $U_2 \perp_{\cap} V_2$ is also true, and*

$$U \cap V = (U_1 \cap V_1) \oplus (U_1 \cap V_2) \oplus (U_2 \cap V_1) \oplus (U_2 \cap V_2).$$

As a direct corollary of Lemma 2 and Lemma 3, we obtain

Corollary 1. *Let $\mathcal{L}(\mathcal{R}_i \times \mathcal{C} : i = 1, \dots, s)$ and $\mathcal{L}(\mathcal{R} \times \mathcal{C}_j : j = 1, \dots, t)$ be two log-linear models, and suppose that $\mathcal{R} \succ \mathcal{R}_i$ and $\mathcal{C} \succ \mathcal{C}_j$ for all $1 \leq i \leq s$, $1 \leq j \leq t$. Then the intersection of the two models is the log-linear model $\mathcal{L}(\mathcal{R}_i \times \mathcal{C}_j : i = 1, \dots, s, j = 1, \dots, t)$.*

3.2 Decomposability as a log-linear model

In this section, we prove Theorem 1. Recall from (8) the definition of a consecutive partition of $\{1 : n\}$. A consecutive partition, which contains only two neighboring sections is called a *bold section*. That is, the k th bold section, containing the sections $k - 1$ and k , is given by

$$\Phi_k = (\{1 : k - 1\}, \{k\}, \{k + 1 : n\}), \quad 2 \leq k \leq n - 1. \quad (23)$$

We will extend the notation Φ_k to $k = 1$ and $k = n$ for the sake of convenience. The (consecutive) partition which partitions $\{1 : n\}$ into n sets is called the *full partition*:

$$\Psi = (\{1\}, \{2\}, \dots, \{n\}). \quad (24)$$

From the multiplicative form (6), it is straightforward that the L -decomposable exponential family \mathcal{P}_L^+ is the log-linear model

$$\mathcal{P}_L^+ = \mathcal{L}(\Phi_k \times \Psi : 1 \leq k \leq n). \quad (25)$$

That is, the generators are the products of bold sections with the full partition. The reason is that the $0 - 1$ matrix $|\pi_{\Phi_k \times \Psi}|$ is equivalent to the vector of unordered marginals $\{\pi_{\Phi_k}\}$ defined in (9).

A submodel which we will use in the sequel has as generators coarser partitions. A consecutive partition, which contains only one section is called a *thin section*. Denote the k th thin section by

$$\tilde{\Phi}_k = (\{1 : k\}, \{k + 1 : n\}), \quad 1 \leq k \leq n - 1. \quad (26)$$

Notice that $\Phi_1 = \tilde{\Phi}_1$ and $\Phi_n = \tilde{\Phi}_{n-1}$. We extend the notation $\tilde{\Phi}_k$ to $k = n$ for the sake of convenience. The log-linear model

$$\mathcal{P}_{L_S}^+ = \mathcal{L}(\tilde{\Phi}_k \times \Psi : 1 \leq k \leq n - 1) \quad (27)$$

is a submodel of the L -decomposable family, which consists of positive distributions p for which the conditional probability $P(\Pi(|C| + 1) = x \mid \Pi\{1 : |C|\} = C)$ depends on the pair (x, C) only through their union, $C \cup \{x\}$, where, as before, Π is a random permutation with distribution p . We will call these distributions L_S -decomposable, where S stands for “set”, indicating that the choice of the k th element of the random permutation depends only on the set to be formed by the first k elements. We define L'_S -decomposable distributions similarly.

Recall from (19) the definition of the subspace corresponding to a product partition, and for the sake of brevity introduce the notations

$$U^{\Phi_k \times \Psi} = U^k, \quad U^{\tilde{\Phi}_k \times \Psi} = \tilde{U}^k. \quad (28)$$

From the partial ordering of partitions, we get that $\tilde{U}^k \subset U^k$, denote the orthogonal complement of \tilde{U}^k in U^k by F^k . By the same argument, $\tilde{U}^k \subset U^{k+1}$ also holds. This yields that

$$\text{Span}(U^k : 1 \leq k \leq n) = \text{Span}(F^k : 1 \leq k \leq n, \tilde{U}^n).$$

Since $\Phi_1 = \tilde{\Phi}_1$, we get $F^1 = \{\mathbf{0}\}$. In addition, as $\tilde{\Phi}_n$ is the trivial partition, $\tilde{U}^n = \text{Span}(\mathbf{1})$. Thus the subspace belonging to the L -decomposable log-linear model is

$$F = \text{Span}(U^k : 1 \leq k \leq n) = \text{Span}(F^k : 2 \leq k \leq n, \mathbf{1}). \quad (29)$$

In the next lemma, we show that the subspaces on the righthandside of (29) not only span F , but give an orthogonal decomposition. The proof is found in Section 7.

Lemma 5. *The subspaces F^k ($2 \leq k \leq n$) are orthogonal to each other and to the vector $\mathbf{1}$.*

The number of free parameters in the L -decomposable exponential family, which we denote by b_n , is the dimension of F minus one. From the orthogonal decomposition (29) of F it is immediate that

$$b_n = \dim(F) - 1 = \sum_{k=2}^n \dim(F^k) = \sum_{k=2}^n \binom{n}{k} (k-1) = 2^n(n/2 - 1) + 1, \quad (30)$$

where the dimension of F^k is easy to calculate.

The decomposition (29) simplifies the calculations regarding the dimension of the bi-decomposable model as well. Interchanging the role of rows and columns, we see that the L' -decomposable loglinear family is

$$\mathcal{P}_{L'}^+ = \mathcal{L}(\Psi \times \Phi_k : 1 \leq k \leq n). \quad (31)$$

Define the (k, ℓ) th *bold cross-section* as

$$\mathcal{H}_{k\ell} = \Phi_k \times \Phi_\ell, \quad (32)$$

where the component (row and column) partitions are bold sections defined in (23). By Corollary 1, the bi-decomposable log-linear model is given by

$$\mathcal{P}_b^+ = \mathcal{L}(\mathcal{H}_{k\ell} : 1 \leq k, \ell \leq n). \quad (33)$$

In order to find the dimension of H , define the subspaces $V^\ell, \tilde{V}^\ell, G^\ell$ in the L' -decomposable model just as we defined U^k, \tilde{U}^k, F^k in the L -decomposable model. Using the notation introduced in (19),

$$U^{\Psi \times \Phi_\ell} = V^\ell, \quad U^{\Psi \times \tilde{\Phi}_\ell} = \tilde{V}^\ell. \quad (34)$$

Applying Lemma 5, the subspace corresponding to the L' -decomposable model can be written as

$$G = \oplus_{\ell=2}^n G^\ell \oplus \text{Span}(\mathbf{1}). \quad (35)$$

By Lemma 2, for any pair

$$U \in \{U^k, \tilde{U}^k : 2 \leq k \leq n\}, \quad V \in \{V^\ell, \tilde{V}^\ell : 2 \leq \ell \leq n\},$$

$U \perp_\cap V$, since these subspaces correspond to product partitions, where in the U -partitions, the column-partition is as fine as possible, and in the V -partitions, the row-partition is as fine as possible. By Lemma 4, we get $F^k \perp_\cap G^\ell$ for all k, ℓ . By Lemma 3, the space $H = F \cap G$ corresponding to bi-decomposable distributions has the orthogonal decomposition

$$H = \oplus_{2 \leq k, \ell \leq n} (F^k \cap G^\ell) \oplus \mathbf{1}. \quad (36)$$

It remains to find the dimension and a basis of $F^k \cap G^\ell$. For the time being, fix k and ℓ . By Lemma 2, the subspace corresponding to the (k, ℓ) th bold cross-section $\mathcal{H}_{k\ell}$ is just $U^k \cap V^\ell$. Observe that $F^k \cap G^\ell$ consists of exactly those vectors of the space $U^k \cap V^\ell$, which are orthogonal to both \tilde{U}^k and \tilde{V}^ℓ . Recall that the π th coordinate of a vector in $U^k \cap V^\ell$ depends only on its marginal $|\pi_{\mathcal{H}_{k\ell}}| = (t_{ij})_{1 \leq i, j \leq 3}$ defined by (18), that is the number of rooks π places in the nine parts into which the (k, ℓ) th bold cross-section divides the chessboard.

As the nine elements of the matrix $|\pi_{\mathcal{H}_{k\ell}}|$ must satisfy row-sum and column-sum constraints, the vector is determined by its coordinates t_{ij} for $i, j = 1, 2$. Furthermore, all of t_{12}, t_{21}, t_{22} can be either zero or one. We will specify the marginal $|\pi_{\mathcal{H}_{k\ell}}|$ by two coordinates: $a = t_{11} + t_{12} + t_{21} + t_{22}$ and q , where q codes the placement of the rooks in the middle row and column of the 3×3 partition. Our coding is as follows. In the k th row and ℓ th column, there is either one rook in the intersection of the row and the column (code 5), or there are two rooks, one on a horizontal, and one on a vertical arm of the cross. In this latter case, the two occupied arms point towards a plane-quarter, and we use the usual numbering of the plane-quarters as coding (one or two arms of the cross may be missing, but this does not cause any problems). That is, for fixed k, ℓ ,

$$a^{k\ell}(\pi) = |\{i : 1 \leq i \leq k, 1 \leq \pi(i) \leq \ell\}|, \quad (37)$$

and

$$q^{k\ell}(\pi) = \begin{cases} 1 & \text{if } \pi(k) > \ell, \quad \pi^{-1}(\ell) < k \\ 2 & \text{if } \pi(k) < \ell, \quad \pi^{-1}(\ell) < k \\ 3 & \text{if } \pi(k) < \ell, \quad \pi^{-1}(\ell) > k \\ 4 & \text{if } \pi(k) > \ell, \quad \pi^{-1}(\ell) > k \\ 5 & \text{if } \pi(k) = \ell. \end{cases} \quad (38)$$

Since the π th coordinate of a vector in $U^k \cap V^\ell$ depends only on its marginal $|\pi_{\mathcal{H}_{k\ell}}|$, a basis of $U^k \cap V^\ell$ is given by the indicator vectors of all the possible values of this marginal. Therefore, for each a, q , we define this indicator vector

$$\rho_{aq}^{k\ell}(\pi) = \chi\{a^{k\ell}(\pi) = a, q^{k\ell}(\pi) = q\}. \quad (39)$$

Of course, for many pairs a, q , these are zero vectors. We now determine all cases when $\rho_{aq}^{k\ell}$ is not identically zero. First, we need

$$\max(0, k + \ell - n) \leq a \leq \min(k, \ell), \quad (40)$$

as there must be a non-negative number of rooks in each rectangle of the board. q can usually be anything from 1 to 5, except

$$\begin{aligned} a = 0 & \Rightarrow q = 4, \\ a = 1 & \Rightarrow q \in \{1, 3, 4, 5\}, \\ a = k < j & \Rightarrow q \in \{2, 3, 5\}, \\ a = j < k & \Rightarrow q \in \{1, 2, 5\}, \\ a = j = k & \Rightarrow q \in \{2, 5\}. \end{aligned} \quad (41)$$

We call the pairs a, q satisfying (40) and (41) non-trivial pairs. After all this preparation, we are ready for the proof of Theorem 1.

Proof of Theorem 1. We have to show (15). The number of free parameters of the bi-decomposable log-linear model is $\dim(H) - 1$, since only those vectors v in H are allowed for which $p = e^v$ is a probability distribution. By (36), we only need to determine the dimension of each subspace $F^k \cap G^\ell$, which consists of the vectors of $U^k \cap V^\ell$, which are orthogonal to both \tilde{U}^k and \tilde{V}^ℓ .

Let $u = \sum_{a,q} c_{aq} \rho_{aq}^{k\ell}$ be an arbitrary vector in $U^k \cap V^\ell$, we find when it is orthogonal to \tilde{U}^k and \tilde{V}^ℓ . First take a vector $v(\pi) = \chi(\pi\{1 : k\} = C)$ in the basis of \tilde{U}^k , and introduce the notation $|C \cap \{1 : \ell\}| = a$. With $h = (k-1)!(n-k)!$, the scalar product is calculated as

$$(u, v) = \begin{cases} c_{a1}(k-a)h + c_{a2}(a-1)h + c_{a5}h & \text{if } \ell \in C \\ c_{a3}ah + c_{a4}(k-a)h & \text{if } \ell \notin C \end{cases}$$

Similarly, if $v(\pi) = \chi(\pi^{-1}\{1 : \ell\} = D)$ is a basis vector of \tilde{V}^ℓ , $|D \cap \{1 : k\}| = a$, and $g = (\ell-1)!(n-\ell)!$, then

$$(u, v) = \begin{cases} c_{a3}(\ell-a)g + c_{a2}(a-1)g + c_{a5}g & \text{if } k \in D \\ c_{a1}ag + c_{a4}(\ell-a)g & \text{if } k \notin D \end{cases}$$

Thus $F^k \cap G^\ell$ consists of the linear combinations of those vectors $\sum_{q=1}^5 c_{aq} \rho_{aq}^{k\ell}$ for which the above four linear combinations of the coefficients c_{aq} are zero. Of the four constraints on

the coefficients, only three are linearly independent, so in most cases there are two linearly independent solutions for the five coefficients. The cases $a = 0, 1, \min(k, \ell)$ must be treated separately, it is readily seen that in the case $a = 0$, the only solution is zero, while in the cases $a = 1, \min(k, \ell)$ there is one non-zero solution. Let $\Delta_a^{k\ell}$ denote the number of linearly independent solutions, that is $\Delta_a^{k\ell}$ is either zero, one or two. The following vectors form an orthogonal basis of $F^k \cap G^\ell$ (with the exception that some vectors may be $\mathbf{0}$):

$$\begin{aligned}\mu_{a1}^{k\ell} &= -\rho_{a2}^{k\ell} + (a-1)\rho_{a5}^{k\ell} \\ \mu_{a2}^{k\ell} &= -(\ell-a)a\rho_{a1}^{k\ell} + (k-a)(\ell-a)\rho_{a2}^{k\ell} - (k-a)a\rho_{a3}^{k\ell} + \\ &\quad + a^2\rho_{a4}^{k\ell} + (k-a)(\ell-a)\rho_{a5}^{k\ell}\end{aligned}\tag{42}$$

Finally, since

$$\sum_{k,\ell} \dim(F^k \cap G^\ell) = \sum_{i \geq 1} |\{(k, \ell) : \dim(F^k \cap G^\ell) \geq i\}|,$$

to finish the proof of the theorem, it suffices to show that for $1 \leq i \leq n$

$$|\{(k, \ell) : \dim(F^k \cap G^\ell) \geq i\}| = (n-i)^2.$$

To this end, let us find those k, ℓ , for which $\dim(F^k \cap G^\ell) \geq 2j+2$. This happens if among the quantities $\Delta_a^{k\ell}$ there are either two 1's and at least j 2's, or one 1 and at least $(j+1)$ 2's.

The first of these cases occurs when $\ell + k \leq n+1$ and $\min\{k, \ell\} \geq j+2$, while the second case occurs when $\ell + k \geq n+2$ and $\max\{k, \ell\} \leq n-j-1$. But if $\ell + k \leq n+1$ and $k, \ell \geq j+2$, then $k, \ell \leq n-j-1$ also holds. Similarly, if $\ell + k \geq n+2$ and $k, \ell \leq n-j-1$, then at the same time $k, \ell \geq j+3 > j+2$. Therefore, $\dim(F^k \cap G^\ell) \geq 2j+2$ holds if and only if $j+2 \leq k, \ell \leq n-j-1$, and there are $[n-(2j+2)]^2$ such pairs.

Let us find those k, ℓ , for which $\dim(F^k \cap G^\ell) \geq 2j+1$. This happens if among the quantities $\Delta_a^{k\ell}$ there are either two 1's and at least j 2's, or one 1 and at least j 2's.

The first of these cases occurs when $\ell + k \leq n+1$ and $\min\{k, \ell\} \geq j+2$, while the second case occurs when $\ell + k \geq n+2$ and $\max\{k, \ell\} \leq n-j$. But if $\ell + k \leq n+1$ and $k, \ell \geq j+2$, then $k, \ell \leq n-j-1 < n-j$ also holds. Similarly, if $\ell + k \geq n+2$ and $k, \ell \leq n-j$, then at the same time $k, \ell \geq j+2$. Therefore, $\dim(F^k \cap G^\ell) \geq 2j+1$ holds if and only if $j+2 \leq k, \ell \leq n-j$, and there are $[n-(2j+1)]^2$ such pairs. \square

Remark 1. In (42), we found an orthogonal basis $\{\mathbf{1}, \mu_{ai}^{k\ell}\}$ of the space H . This orthogonality is convenient for finding the parameters corresponding to a bi-decomposable distribution. There exists a basis consisting of indicator vectors as well, as follows. Denote for any k, ℓ, a $\nu_a^{k\ell} = \sum_{q=1}^5 \rho_{aq}^{k\ell}$, where $\rho_{aq}^{k\ell}$ was defined in (39). That is, $\nu_a^{k\ell}$ is the indicator

vector of the event that there are exactly a rooks in the upper left $k \times \ell$ rectangle of the chessboard. The following vectors, together with $\mathbf{1}$, form a basis of H :

$$\begin{aligned} \nu_a^{k\ell} : & \quad 1 \leq k, \ell \leq n-1, \max(0, k+\ell-n) < a \leq \min(k, \ell), \\ \rho_{a5}^{k\ell} : & \quad 1 \leq k, \ell \leq n-1, \max(1, k+\ell-n) < a \leq \min(k, \ell). \end{aligned} \quad (43)$$

This statement can be proved by induction, we omit the somewhat lengthy calculations.

Remark 2. Notice that the vectors $\nu_a^{k\ell}$, together with $\mathbf{1}$ form the basis of the subspace associated with those positive distributions, called *big-decomposable* distributions, which belong to the intersection of the L_S - and L'_S -decomposable models. By Corollary 1, this is again a log-linear model, with generating product partitions

$$\tilde{\mathcal{H}}_{k\ell} = \tilde{\Phi}_k \times \tilde{\Phi}_\ell, \quad (44)$$

which we call the *thin* (k, ℓ) *cross-sections*, dividing the chessboard into four rectangles. The component partitions were defined in (26). Notice that in this case, with k, ℓ fixed, $\nu_a^{k\ell}$ (a takes on all its possible values) is an orthogonal basis of the subspace corresponding to $\tilde{\mathcal{H}}_{k\ell}$. These subspaces are “almost linearly independent” in the sense that the only linear dependence is that they all contain the vector $\mathbf{1}$. From this it follows that the number of parameters in this model is

$$e_n = \sum_{j=0}^{\lfloor (n-1)/2 \rfloor} (n-2j-1)^2. \quad (45)$$

The vectors $\rho_{a5}^{k\ell}$ represent the “difference” between the bi-decomposable and the big-decomposable distributions.

Remark 3. We have calculated the number of free parameters in the L -decomposable, bi-decomposable, big-decomposable models. For the sake of completeness we mention that the number of free parameters in the remaining L_S -decomposable model is given by

$$c_n = \sum_{i=1}^n \left[\binom{n}{i} - 1 \right] = 2^n - n - 1. \quad (46)$$

3.3 Maximum likelihood estimation

As we have seen in the previous sections, the positive bi-decomposable distributions \mathcal{P}_b^+ on S_n form an exponential family with d_n parameters.

Denote by Π_1, \dots, Π_m a sample taken from a positive bi-decomposable distribution, and let $r(\pi)$ stand for the relative frequency of the permutation π in the sample. The maximum likelihood estimate of the true distribution, or equivalently, of its parameters, does not appear to have an explicit form in general, the likelihood function has to be

maximized by numerical methods. The iterative proportional fitting procedure (IPFP), used in the theory of log-linear models, is one option. This algorithm converges to the maximum likelihood estimate, if it exists. We describe briefly the implementation of this algorithm in our setting.

The generators of the bi-decomposable log-linear model are the bold cross-sections $\mathcal{H}_{k\ell}$ in (32), which define the marginal functions $|\pi_{\mathcal{H}_{k\ell}}|$ given by (18). The maximum likelihood estimate is a distribution $p^* \in \mathcal{P}_b^+$ such that the distributions of the marginals $|\Pi_{\mathcal{H}_{k\ell}}|$ under p^* are the same as under the empirical distribution r . There is at most one such p^* in \mathcal{P}_b^+ . In some cases, the maximum likelihood estimate does not exist, because no distribution in \mathcal{P}_b^+ gives the same distribution of the marginals as the empirical distribution (we say that the sample contains structural zeros). In these cases, a suitable p^* can only be found in the closure $cl(\mathcal{P}_b^+)$. Numerical studies indicate that $cl(\mathcal{P}_b^+) = \mathcal{P}_b$.

The IPFP algorithm proceeds by cyclically fitting the distributions of the individual marginals $|\Pi_{\mathcal{H}_{k\ell}}|$ to that observed in the sample. It converges to the unique element in $cl(\mathcal{P}_b^+)$ which agrees with the empirical distribution in all marginals. Starting from an arbitrary $p^1 \in \mathcal{P}_b^+$ (say the uniform distribution), the n th iteration step calculates

$$p^{n+1}(\pi) = \frac{\sum_{\sigma: |\sigma_{\mathcal{H}_{k\ell}}| = |\pi_{\mathcal{H}_{k\ell}}|} r(\sigma)}{\sum_{\sigma: |\sigma_{\mathcal{H}_{k\ell}}| = |\pi_{\mathcal{H}_{k\ell}}|} p^n(\sigma)} p^n(\pi), \quad (47)$$

where the pair (k, ℓ) runs cyclically over all possible values.

Remark 4. By Remark 2, maximum likelihood estimation in the big-decomposable model proceeds in an analogous way, namely by running the IPFP algorithm with $|\pi_{\tilde{\mathcal{H}}_{k\ell}}|$ of (44) instead of $|\pi_{\mathcal{H}_{k\ell}}|$, i.e. we use the thin cross-sections instead of the bold ones.

Remark 5. In \mathcal{P}_L and in $\mathcal{P}_{L'}$, the maximum likelihood estimate can be given explicitly. For example, the L -decomposable model is parametrized by the conditional probabilities (4). The maximum likelihood estimate of these conditional probabilities is given by the corresponding conditional probabilities under the empirical distribution.

Numerical studies indicate that the maximum likelihood estimate in the family $cl(\mathcal{P}_b^+)$ can also be obtained by iteratively calculating the maximum likelihood projections on the component spaces \mathcal{P}_L and $\mathcal{P}_{L'}$.

4 Examples

In this section, we collect some models from the literature, which are decomposable in at least one way. These models are submodels of the “free” decomposable models, since they place specific constraints on the parameters.

Example 1 (Order statistics models). Consider an experiment in which people are asked to rank sounds according to their loudness, say in increasing order. One might suppose

that the actual perception of each stimulus is a random variable, whose relative ordering determines the person's ordering of the sounds. This example was studied by Thurstone [20], and later by Daniels [5]. If the random variable associated to the i th sound is X_i with continuous distribution F_i , then the resulting distribution on the orderings is

$$p(\pi) = P(X_{\pi(1)} < \cdots < X_{\pi(n)}).$$

The assumption that X_i are independent leads to the so called *order statistics models*. If the distributions form a location family $F_i(x) = F(x - \mu_i)$, the model is called *Thurstone model*. A well-studied case is when $F(x) = 1 - \exp(-\exp x)$ is the Gumbel distribution. This is called *Luce model*, and is equivalent to the model derived by Luce on the basis on his ranking postulate and Choice Axiom in [15]. This model is L -decomposable, with canonical decomposition

$$\Lambda(x, C) = \frac{\theta_x}{\sum_{y \notin C} \theta_y},$$

where θ_y are arbitrary positive parameters (with sum equal to 1) associated with the objects. However, the model is not L' -decomposable for arbitrary θ .

Example 2 (Paired comparisons models). A model suggested by Babington Smith [1] creates the ordering of the n objects by making every possible paired comparison independently of each other. The result of such a tournament can be represented by a directed graph: if the graph contains no directed circle, then it corresponds to a unique ordering of the objects. Conditioning on the event that the graph is circle-free, we get

$$p(\pi) = c(\theta) \prod_{i < j} \theta_{\pi(i)\pi(j)},$$

where θ_{xy} is the probability that object x is preferred to object y in a paired comparison. This model is L -decomposable, with

$$\Lambda(x, C) = \prod_{y \in C} \theta_{yx}.$$

However, the model is not L' -decomposable for arbitrary parameters.

Example 3 (Mallows-Bradley-Terry model). A special case of the paired comparison model is given by $\theta_{xy} = \frac{\alpha_x}{\alpha_x + \alpha_y}$. This form of the paired comparison probabilities was suggested by Bradley and Terry [2], and Mallows [16] suggested using these probabilities in the paired comparison model. The resulting distribution on the orderings is given by

$$p(\pi) = c(\alpha) \prod_{i=1}^n \alpha_{\pi(i)}^{n-i} = c(\alpha) \prod_{j=1}^n \alpha_j^{n-\pi^{-1}(j)},$$

which is bi-decomposable.

Example 4 (Multistage ranking model). This model, investigated by Fligner and Verducci [11] supposes the candidates are numbered from 1 to n . The ranking takes place stepwise. In the k th step, the best $k - 1$ ranks are already given out. The k th best candidate is then chosen from the remaining ones, but only the relative order of the remaining candidates is taken into account. In particular, if the remaining candidates are $j_1 < \dots < j_{n-k+1}$, then choose j_i with probability $\theta(i, k)$, where $\theta(i, k)$ are parameters satisfying $\sum_{i=1}^{n-k+1} \theta(i, k) = 1$. It is easily seen that this model is L -decomposable with

$$\Lambda(x, C) = \theta(|\overline{C} \cap \{1 : x\}|, |C| + 1).$$

The model is also L' -decomposable.

Example 5 (Repeated insertion model). This model, studied by Doignon, Pekeč and Regenwetter [9] assumes that the ordering is created by considering the candidates one after the other (according to their fixed numbering), and inserting the current candidate into the order already formed by the previous ones. More specifically, for each k , we have insertion probabilities $\theta(i, k)$, $i = 1, \dots, k$ with sum 1. For the k th candidate, there are k possible places where he or she can be inserted into the order of the first $k - 1$ candidates: insert him or her between the $(i - 1)$ st and i th with probability $\theta(i, k)$. This model is a “dual” of the multistage ranking model, in the sense that it can be described similarly to it, by interchanging “ranks” and “candidates” (but not in the sense that the resulting permutations are each other’s inverse). It is L -decomposable with

$$\Lambda(x, C) = \theta(|C \cap \{1 : x\}| + 1, x),$$

and it is also L' -decomposable.

Example 6 (Quasi-independence log-linear model). Let θ_{ij} , $1 \leq i, j \leq n$, be the elements of an arbitrary doubly stochastic matrix, and

$$p(\pi) = c(\theta) \prod_{i=1}^n \theta_{i\pi(i)} = c(\theta) \prod_{j=1}^n \theta_{\pi^{-1}(j)j}.$$

This distribution is by the above equation bi-decomposable. Writing it in the log-linear form

$$\log p(\pi) = \alpha^{(0)} + \alpha_{\pi(1)}^{(1)} + \alpha_{\pi(2)}^{(2)} + \dots + \alpha_{\pi(n)}^{(n)}, \quad (48)$$

it states the quasi-independence of the variables $\pi(i)$, $1 \leq i \leq n$.

It is easy to see that for $n > 3$ the random permutations belonging to the quasi-independence model have the following property. For any partition $\mathcal{Z} = (Z_1, Z_2)$ of size 2, as in (16), with $|Z_1| = 2$, the ordered marginals Π_{Z_1} and Π_{Z_2} are conditionally independent, given the unordered marginals $\{\Pi_{\mathcal{Z}}\}$. This property is a generalization of

L -decomposability, and we can prove that only the quasi-independent distributions possess it. Moreover, for these distributions, a similar property is satisfied with arbitrary partitions \mathcal{Z} .

5 Invariance under relabellings

As we have noted already, decomposability of a random pairing between sets A and B depends on how we label the elements of the two sets. Suppose that a labelling on both sets is fixed, and the random pairing function $\Pi : A \rightarrow B$ with these labellings becomes $\Pi_{orig} : \{1 : n\} \rightarrow \{1 : n\}$. Suppose that we relabel the set A according to the permutation $\sigma \in S_n$, that is object with original label i receives the new label $\sigma(i)$. Similarly, relabel the set B according to $\rho \in S_n$. Denote the random pairing function $\Pi : A \rightarrow B$ with these new labellings $\Pi_{new} : \{1 : n\} \rightarrow \{1 : n\}$. If with the original labelling, the pair of $i \in A$ is $\pi(i) \in B$, then with the new labelling, the pair of $\sigma(i) \in A$ is $\rho\pi(i) \in B$. Therefore, for the distributions p_{orig} and p_{new} ,

$$p_{orig}(\pi) = P(\Pi_{orig} = \pi) = P(\Pi_{new} = \rho\pi\sigma^{-1}) = p_{new}(\rho\pi\sigma^{-1}). \quad (49)$$

In this section we investigate whether L -decomposability is preserved after such relabellings or not.

Definition 4. Let $\phi : S_n \rightarrow S_n$ be a one to one mapping. Then for any distribution p , define $p_\phi(\pi) = p(\phi(\pi))$. We say that the family \mathcal{P} of distributions is *invariant* under ϕ , if

$$\mathcal{P}_\phi = \{p_\phi : p \in \mathcal{P}\} \subset \mathcal{P}.$$

Let us introduce some notation. For any $\sigma \in S_n$, let $\phi_{\circ\sigma} : \pi \mapsto \pi\sigma$, $\phi_{\sigma\circ} : \pi \mapsto \sigma\pi$ be the right and left multiplications by σ . Denote by $\sigma_{(12)}$ the permutation which exchanges 1 and 2 only, and by σ_r the reversing permutation which maps k to $n + 1 - k$.

In the ranking situation (ii) described in the Introduction, a model for a random ranking is called *label-invariant*, if it is invariant under relabellings of the objects. It is called *reversible*, if it is invariant under reversing of the ranks. The concepts of label-invariance and reversibility were studied for some wide classes of ranking models in [4].

Theorem 2. \mathcal{P}_L is invariant under left multiplications, and under the group of right multiplications generated by σ_r and $\sigma_{(12)}$ (for $n \geq 4$, this group contains eight right multiplications, including the identity). The family is not invariant under any other right multiplication.

Proof. Invariance under left-multiplications follows e.g. from Property 3 in Definition 1, as well as invariance under right multiplication by σ_r , since the Markov property is reversible.

Invariance under right multiplication by $\sigma_{(12)}$ can be checked directly using Property 1 in Definition 1.

To show that the family is not invariant under other right multiplications, we prove that for all other permutations σ there exists a positive bi-decomposable distribution p , such that $p_{\circ\sigma}$ is not L -decomposable. The group of right multiplications generated by $\phi_{\circ\sigma_r}, \phi_{\circ\sigma_{(12)}}$ are the multiplications by the following permutations:

$$id, \sigma_r, \sigma_{(12)}, \sigma_r\sigma_{(12)}, \sigma_r\sigma_{(12)}\sigma_r, \sigma_{(12)}\sigma_r, \sigma_r\sigma_{(12)}\sigma_r\sigma_{(12)}, \sigma_{(12)}\sigma_r\sigma_{(12)} \quad (50)$$

We will use the following property: let p be L -decomposable. Suppose that the probability of π_{11} and π_{22} is positive, and a is such that $\pi_{11}\{1 : a\} = \pi_{22}\{1 : a\}$. Define the “crossover” permutations:

$$\pi_{12}(k) = \begin{cases} \pi_{11}(k) & \text{if } k \leq a \\ \pi_{22}(k) & \text{if } k > a \end{cases}, \quad \pi_{21}(k) = \begin{cases} \pi_{22}(k) & \text{if } k \leq a \\ \pi_{11}(k) & \text{if } k > a \end{cases}.$$

Then

$$\frac{p(\pi_{11})}{p(\pi_{12})} = \frac{p(\pi_{21})}{p(\pi_{22})}. \quad (51)$$

If σ not a member of the permutations in (50), then neither is its inverse, and there exists an $2 \leq a \leq n - 2$, such that

$$\sigma^{-1}\{1 : a\} \neq \{1 : a\}, \{n - a + 1 : n\}.$$

Let a be such a number. Therefore there exist $c, e \in \{1 : a\}$ and $d, f \notin \{1 : a\}$, for which

$$c^* = \sigma^{-1}(c) > \sigma^{-1}(d) = d^*, \quad e^* = \sigma^{-1}(e) < \sigma^{-1}(f) = f^*.$$

For the numbers α, β, γ we say that α separates β and γ , if $\beta < \alpha < \gamma$ or $\beta > \alpha > \gamma$. Now, if $d^* \geq f^*$, then d^* (and f^* as well) separates c^* and e^* . If $d^* < f^*$, then either one of them separates c^* and e^* , or c^* (and e^* as well) separates d^* and f^* . Therefore, one of the following two cases holds:

1. $\exists c, e \in \{1 : a\}, \quad d \notin \{1 : a\} : \quad d^*$ separates c^*, e^*
2. $\exists c \in \{1 : a\}, \quad d, f \notin \{1 : a\} : \quad c^*$ separates d^*, f^*

The two cases can be treated in the same way. Let us deal with the first one! Let $f \notin \{1 : a\}, f \neq d$ be arbitrary, with $f^* = \sigma^{-1}(f)$. Recall (39), and let $p = c(d^*) \exp\{\rho_{d^*5}^{d^*}\}$, this is a positive bi-decomposable distribution. Let $\pi_{11} = \sigma^{-1}$, from which we obtain π_{22} by exchanging two pairs:

$$\pi_{22}(c) = e^*, \pi_{22}(e) = c^*, \pi_{22}(d) = f^*, \pi_{22}(f) = d^*.$$

Denote by π_{12} and π_{21} the crossover permutations. For these four permutations, $p_{\circ\sigma}$ does not satisfy (51). On the one hand, multiplying π_{11} by σ from the right, we get the identity permutation, for which $\rho_{d^*5}^{d^*d^*} = 1$. On the other hand, for both $\pi_{12}\sigma$ and $\pi_{21}\sigma$, $\rho_{d^*5}^{d^*d^*} = 0$, since for the first, d^* is not a fixed point, and for the second, there is an element greater than d^* among the first d^* elements. This completes the proof. \square

6 Discussion and application

In this paper, we introduced log-linear models for random permutations, whose generators are product partitions of the chessboard. Examples are the L -decomposable, L_S -decomposable, bi-decomposable, and big-decomposable models. In all of these cases, we determined the number of parameters. We showed how to calculate the maximum likelihood estimate of the continuous parameters either directly (for the L -decomposable model) or by the iterative proportional fitting algorithm (in the other models). The natural order(s) implied by the models on the set(s) is either known, or it also has to be estimated. We studied the extent to which this order can be determined in the L -decomposable and the bi-decomposable models. There are many other statistical questions of interest, which we did not address in this paper. Another theoretically, and perhaps also practically important question is the characterization of decomposable distributions, if we do not restrict ourselves to the strictly positive case.

Finally, we fit our models to one of the most investigated ranking data in the literature, the 1980 election of the American Psychological Association (APA). This organization elects a president each year by asking its members to rank five candidates. In 1980, 5738 complete rankings were cast. APA chooses the winner by the Hare system. See Fishburn [10] for a review of the advantages and disadvantages of this system.

Analyses of these data can be found, among others, in Chung and Marden [3], Diaconis [7], McCullagh [18], and Stern [19]. One characteristic feature of the data is that the members of the association can be divided into three distinct groups: the research psychologists (candidates 1 and 3 belong here), clinical psychologists (their candidates are 4 and 5), and the community psychologists, to whom 2 belongs. The first two groups represent the majority of the members. Not surprisingly, analysis shows that each group tends to prefer its own candidates.

Chung and Marden [3] fit orthogonal contrast models to the data. Diaconis [7] uses this dataset to illustrate the method of spectral analysis of ranked data, with many pointers to literature. McCullagh [18] fits log-linear models based on inversions to the data. We emphasize here that we do not attempt to provide a thorough analysis of the APA data, our aim is merely to illustrate the fit of our models on a real dataset. We used the ordering

data, Table 1 shows the maximum of the log-likelihood function (L) and the chi-square value of the goodness of fit, with the degrees of freedom in parentheses for all models. In the last column, we gave the standardized statistic $U = (GOF - df)/\sqrt{df}$. In all cases, where applicable, the results appearing in Table 1 correspond to the best right/left relabelling of the original data, which we found by an exhaustive search. As expected, the order on the ranks indicated by the models coincides with the natural order from best to worst. All models agree on the natural order of the candidates as well, 4 is in the middle, with $\{1, 3\}$ and $\{2, 5\}$ on its two sides (notice that there are eight permutations fitting this pattern, as stated in Theorem 2). The best fit is provided by the L -decomposable model. The results indicate that the rankings violate decomposability (i.e. conditional independence relations) more than the orderings. There are at least two ways to find better fitting models. Firstly, we could check which conditional independences do not hold. Models which assume decomposability at only some row-sections and some column-sections also fit into the log-linear setting described in this paper. While a general expression for the number of free parameters in these wider models is probably intractable, it can be calculated numerically in any particular case. Secondly, one could try to reduce the number of parameters by selecting a log-linear model with fewer generating partitions.

Table 1: Fit of log-linear models to APA ordering data

Model	L	GOF (df)	U
saturated	-26612	-	
L -decomposable	-26661	98.9 (70)	3.45
L' -decomposable	-26674	126.5 (70)	6.75
L_S -decomposable	-26684	144.8 (93)	5.37
L'_S -decomposable	-26697	171.7 (93)	8.16
bi-decomposable	-26687	151.8 (89)	6.66
bi $_S$ -decomposable	-26701	180.1 (99)	8.15
uniform	-27470	2183.0 (119)	

7 Proofs

Proof of Proposition 3. If the stated conditional independences hold, the random permutation is clearly bi-decomposable. For the other direction, suppose Π is bi-decomposable. By L -decomposability, given $\{\Pi_{\underline{\kappa}}\}$, $\Pi_{\underline{\kappa}_i}$ are conditionally independent. Conditioning on $\{\Pi_{\underline{\lambda}}^{-1}\}$ as well does not ruin this independence, since the additional condition restricts the values of the $\Pi_{\underline{\kappa}_i}$ one by one. Thus we proved that $\Pi_{\underline{\kappa}_i}$ are conditionally independent, given $\{\Pi_{\underline{\kappa}}\}$ and $\{\Pi_{\underline{\lambda}}^{-1}\}$, and using L' -decomposability, the same is true for $\Pi_{\underline{\lambda}_j}^{-1}$. Denote a

condition by $E = \{\{\Pi_{\underline{\kappa}}\} = u, \{\Pi_{\underline{\lambda}}^{-1}\} = v\}$, then

$$P(\Pi = \pi | E) = \prod_{i=1}^s P(\Pi_{\underline{\kappa}_i} = \pi_{\underline{\kappa}_i} | E),$$

and since

$$\Pi_{\underline{\kappa}_i} = (\Pi_{\underline{\kappa}_i \times \underline{\lambda}_j} : 1 \leq j \leq t),$$

where $\Pi_{\underline{\kappa}_i \times \underline{\lambda}_j}$ is a function of $\Pi_{\underline{\lambda}_j}^{-1}$, also

$$P(\Pi_{\underline{\kappa}_i} = \pi_{\underline{\kappa}_i} | E) = \prod_{j=1}^t P(\Pi_{\underline{\kappa}_i \times \underline{\lambda}_j} = \pi_{\underline{\kappa}_i \times \underline{\lambda}_j} | E),$$

which proves the lemma. \square

Proof of Lemma 1. Since $L_2(\mathcal{D}_1 \cap \mathcal{D}_2) = L_2(\mathcal{D}_1) \cap L_2(\mathcal{D}_2)$, the spaces $L_2(\mathcal{D}_1)$ és $L_2(\mathcal{D}_2)$ intersect orthogonally if and only if for any $f \in L_2(\mathcal{D}_1)$, $g \in L_2(\mathcal{D}_2)$,

$$E([f - E(f | \mathcal{D}_1 \cap \mathcal{D}_2)][g - E(g | \mathcal{D}_1 \cap \mathcal{D}_2)]) = 0.$$

If the conditional independence relation holds, then the following stronger equality holds:

$$E([f - E(f | \mathcal{D}_1 \cap \mathcal{D}_2)][g - E(g | \mathcal{D}_1 \cap \mathcal{D}_2)] | \mathcal{D}_1 \cap \mathcal{D}_2) = 0.$$

In the other direction, if the spaces intersect orthogonally, then let $E_1 \in \mathcal{D}_1$, $E_2 \in \mathcal{D}_2$, and denote by C the event that

$$P(E_1 \cap E_2 | \mathcal{D}_1 \cap \mathcal{D}_2) - P(E_1 | \mathcal{D}_1 \cap \mathcal{D}_2)P(E_2 | \mathcal{D}_1 \cap \mathcal{D}_2) > 0.$$

With $f = \chi(E_1)\chi(C)$ and $g = \chi(E_2)\chi(C)$,

$$\begin{aligned} E([f - E(f | \mathcal{D}_1 \cap \mathcal{D}_2)][g - E(g | \mathcal{D}_1 \cap \mathcal{D}_2)]) &= \\ E(E(fg | \mathcal{D}_1 \cap \mathcal{D}_2) - E(f | \mathcal{D}_1 \cap \mathcal{D}_2)E(g | \mathcal{D}_1 \cap \mathcal{D}_2)) &= \\ = E(\chi(C)[P(E_1 \cap E_2 | \mathcal{D}_1 \cap \mathcal{D}_2) - P(E_1 | \mathcal{D}_1 \cap \mathcal{D}_2)P(E_2 | \mathcal{D}_1 \cap \mathcal{D}_2)]) &= 0. \end{aligned}$$

This is possible only if $P(E_1 \cap E_2 | \mathcal{D}_1 \cap \mathcal{D}_2) - P(E_1 | \mathcal{D}_1 \cap \mathcal{D}_2)P(E_2 | \mathcal{D}_1 \cap \mathcal{D}_2) \leq 0$ with probability 1. The reverse inequality is obtained similarly, thus E_1 and E_2 are conditionally independent, given $\mathcal{D}_1 \cap \mathcal{D}_2$. \square

Proof of Lemma 2. We apply Lemma 1 to S_n endowed with the uniform distribution. In this case, orthogonal intersection in the L_2 -space is equivalent to orthogonal intersection in $\mathbb{R}^{n!}$. The second statement in (22) holds, because it is easy to check that $\sigma(\mathcal{R}' \times \mathcal{C}) \cap \sigma(\mathcal{R} \times \mathcal{C}') = \sigma(\mathcal{R} \times \mathcal{C})$. Concerning the first one, we have to prove that $|\Pi_{\mathcal{R}' \times \mathcal{C}}|$ and $|\Pi_{\mathcal{R} \times \mathcal{C}'}|$ are conditionally independent, given $|\Pi_{\mathcal{R} \times \mathcal{C}}|$, if Π is a uniformly distributed random permutation, which is again easy to check. \square

Proof of Lemma 3. By supposition, $Pr_{V_j}U_i \subset U_i$ for every i, j , therefore $Pr_{V_j}U \subset U$ for every j , i.e. U intersects each V_j orthogonally. Consequently, $Pr_U V_j \subset V_j$ for every j , which yields $Pr_U V \subset V$, which was to be proved. On the other hand, let $W = \text{Span}(U_i \cap V_j : i \in I, j \in J)$. Then $Pr_{V_j}U_i \subset W$, furthermore $Pr_U V_j = Pr_{V_j}U \subset W$, which leads to $Pr_U V \subset W$. \square

Proof of Lemma 4. We use that U and V intersect orthogonally if and only if the projection operators onto them commute, that is $Pr_U Pr_V = Pr_V Pr_U$. Now

$$\begin{aligned} Pr_{U_2} Pr_{V_2} &= (Pr_U - Pr_{U_1})(Pr_V - Pr_{V_1}) = \\ &= Pr_U Pr_V - Pr_{U_1} Pr_V - Pr_U Pr_{V_1} + Pr_{U_1} Pr_{V_1}, \end{aligned}$$

and by supposition the operators in all four terms commute, yielding $Pr_{V_2} Pr_{U_2}$. The second statement follows from Lemma 3. \square

Proof of Lemma 5. We use again that orthogonality in the L_2 -space is equivalent to orthogonality in \mathbb{R}^n . In this proof, we use the notation

$$\sigma_k = \sigma(\Phi_k \times \Psi), \quad \tilde{\sigma}_k = \sigma(\tilde{\Phi}_k \times \Psi).$$

An element of F^k is a difference $f_1 = f - E(f \mid \tilde{\sigma}_k)$, where f is σ_k -measurable. Orthogonality to $\mathbf{1}$ means that $E(f_1) = 0$. For the other statement, let g_1 be an element of F^j , where $j > k$. It is easy to check that under the uniform distribution, σ_k and σ_j are conditionally independent, given $\tilde{\sigma}_k$. Therefore,

$$E(f_1 g_1) = E[E(f_1 g_1 \mid \tilde{\sigma}_k)] = E[E(f_1 \mid \tilde{\sigma}_k) E(g_1 \mid \tilde{\sigma}_k)] = 0,$$

since $E(f_1 \mid \tilde{\sigma}_k) = 0$ with probability 1. \square

References

- [1] BABINGTON-SMITH, B. (1950). Discussion of Professor Ross' paper. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **12** 153–162.
- [2] BRADLEY, R. A. and TERRY, M. A. (1952). Rank analysis of incomplete block designs I. *Biometrika* **39** 324–345.
- [3] CHUNG, L. and MARDEN, J. I. (1993). Extensions of Mallows' ϕ Model. In *Probability Models and Statistical Analyses for Ranking Data* (Fligner, M.A. and Verducci, J.S, eds.) 108–139. Springer-Verlag, New York.
- [4] CRITCHLOW, D. E., FLIGNER, M. A. and VERDUCCI, J. S. (1991). Probability models on rankings. *J. Math. Psych.* **35** 294–318.

- [5] DANIELS, H. E. (1950) Rank correlation and population models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **12** 171–181.
- [6] DIACONIS, P. (1988). *Group Representations in Probability and Statistics*, Institute of Mathematical Statistics, Hayward, California.
- [7] DIACONIS, P. (1989). A generalization of spectral analysis with application to ranked data. *Ann. Statist.* **17** 949–979.
- [8] DIACONIS, P. and ERIKSSON, N. (2006). Markov bases for noncommutative Fourier analysis of ranked data. *J. Symbolic Comput.* **41** 173–181.
- [9] DOIGNON, J.-P., PEKEČ, A. and REGENWETTER, M. (2004). The repeated insertion model for rankings: missing link between two subset choice models. *Psychometrika* **69** 33–54.
- [10] FISHBURN, P. (1973). *The Theory of Social Choice*, Princeton University Press, Princeton, N. J.
- [11] FLIGNER, M. A. and VERDUCCI, J. S. (1988). Multistage ranking models. *J. Amer. Statist. Assoc.* **83** 892–901.
- [12] FLIGNER, M. A. and VERDUCCI, J. S. (eds.) (1993). *Probability Models and Statistical Analyses for Ranking Data*, Springer-Verlag, New York.
- [13] GEIGER, D., MEEK, C. and STURMFELS, B. (2006). On the toric algebra of graphical models. *Ann. Statist.* **34** 1463–1492.
- [14] LAURITZEN, S. (1996). *Graphical Models*, Clarendon Press, Oxford.
- [15] LUCE, R. D. (1959). *Individual Choice Behavior*, Wiley, New York.
- [16] MALLOWS, C. L. (1957). Non-null ranking models I. *Biometrika* **44** 114–130.
- [17] MARDEN, J. I. (1995). *Analyzing and Modeling Rank Data*, Chapman&Hall, London.
- [18] MCCULLAGH, P. (1993). Permutations and regression models. In *Probability Models and Statistical Analyses for Ranking Data* (Fligner, M.A. and Verducci, J.S, eds.) 196–215. Springer-Verlag, New York.
- [19] STERN, H. (1993). Probability models on rankings and the electoral process. In *Probability Models and Statistical Analyses for Ranking Data* (Fligner, M.A. and Verducci, J.S, eds.) 173–195. Springer-Verlag, New York.
- [20] THURSTONE, L. L. (1927). A law of comparative judgement. *Psychological Review* **34** 273–286.